

KPM 1 and 2 Documentation Annex

Annex A: Formula for participation

$$Participation_{(year,polar)} = \frac{(Number\ of\ New\ Entrants) + (Number\ of\ Previous\ Entrants)}{Total\ Population_{(year,polar)}}$$

Annex B: Adjusted flows in population

Population estimates must be adjusted in order to account for the fact that addresses and postcodes within collected data may be temporary addresses instead of permanent addresses. Temporary addresses may lead to incorrect categorisation of individuals into POLAR4 quintiles. As a result, population estimates must be adjusted to reflect the correct quintile for individuals whose addresses may be inaccurate.

Adjusted population estimates can be calculated by using estimates of higher education students who have moved from a certain POLAR4 quintile to another. Student records from 2009 onwards (Higher Education Statistics Agency (HESA)) contain information both on students' term time and permanent addresses. The postcodes of these addresses can be converted into Middle Super Output Area (MSOA) codes representing geographical areas, which can then be converted into POLAR4 quintiles. Although such records predating 2009 do not exist, the data on student population flows from one quintile to another can act as a model for adjusting the same year's total population flows.

To do so, the number of higher education students in each quintile based on permanent address as well as the subsection of temporary address is counted. Example B1 presents the number of students with permanent addresses in quintile 1, which is equal to 1,000. The breakdown of the number of these people whose term time addresses are in each quintile where their permanent is in quintile 1 is also counted. For instance, while there are actually 1,000 people who permanently live in quintile 1, the records for only 300 of these people correctly reflect this. Addresses of others may be reported as falling under other quintiles, which does not reflect their permanent address. There are 200 people whose temporary address suggests they are part of quintile 2 but whose permanent address suggests quintile 1 instead. This number is added to the population count for quintile 1 and subtracted from quintile 2. The population is shifted to account for the flow in a similar manner for other POLAR4 quintiles.

Example B1:

Permanent quintile	Temporary quintile	Total number of students
Q1	Q1	300
Q1	Q2	200
Q1	Q3	250
Q1	Q4	150
Q1	Q5	100
Q1		1,000

Once the total population is adjusted for the flows between quintiles, the population for each quintile is then used in the calculation for the participation rate by POLAR 4 quintiles.

Annex C: Matching and accounting for historical data

Individualised historical records from the Higher Education Statistics Agency (HESA) exist from 1995 whilst records from the Individualised Learner Record (ILR) exist from 2002. Student records from HESA capture students who attend a higher education institution and are studying for a higher education qualification, while ILR data contains information on students who are studying for a higher education qualification at a further education college (FEC). Matching student data for a particular year with previous historical records occurs through 'fuzzy matching' or matching on personal identifiers such as names, dates of birth, and postcodes. It differs from exact matching because exact matching requires all characteristics to be exactly the same. 'Fuzzy matching' provides an advantage in the sense that it does not eliminate records in higher education data which may have changed over time due to data entry error or missing information.

The fact that records are only available for a limited number of years creates the following issues:

- Firstly, if the number of new entrants for the most recent year (2017) is to be calculated, there would be a large number of years of historic data available for matching purposes with HESA records but a fewer number of years of records available in from ILR records.
- Secondly, and more importantly, if this procedure were to be performed on a previous year, such as 2002, there would be no matches for ILR records and a significantly reduced number of matches for HESA records. Calculating new entrant rates for less recent years would result in reduced number of matches or "found" students due to the lack of historical records. The lower number of students "found"

would be due to the lack of matching, which does not provide an accurate representation of the reality.

In order to correct for these issues, adjustments must be made to account for the number of years in historical records. Matching for the current year (most recent year) is used as a baseline in order to determine the proportion of people found in previous years out of the total population. The method for calculating the said proportions is displayed in example C1.

To begin, each individual student in each particular year's record is listed as either "found", meaning that they were matched with historical records and includes the year in which they first attended higher education, or "new", meaning that they were not. The number of "found" students are then counted for each year, indicating how many students in the current year (2017) previously participated in higher education for each year prior to the current year (C1.1). The counts of "found" students are then added cumulatively from the earliest existing record year to the most recent year (C1.2).

In C1.1, for the base year 2017, 50 students were not found previously, while 7 were found in 2016, 5 in 2015, 5 in 2014, and 3 in 2013. The total count presented in C1.2 shows 70 total students who attended higher education in 2017. This is used to calculate the found rate (C1.4), which is the proportion of students found in a particular year and prior out of the total students who participated in 2017. For example, the found rate for 2015 in this example is 13/70, because 13 is the cumulative number of students found from 2013 to 2015, and 70 is the total number of students. The found rate is then 0.19, which represents the found rate for 2 years back from the base year (C1.3). The found rate obtained for the base year 2017 is then used as the standard for how many matches are expected to be found depending on how many years of historical data is searched for (C1.4). It is used to fill in the missing found rates for when actual historical data may not be available.

Example C1: Calculating found rates for 2017

	Year	2017 (base year)	2016	2015	2014	2013
C1.1	Count of "found" students for each year	50 (not matched with historical records)	7	5	5	3
C1.2	Total count (cumulative from earliest year)	70	20	13	8	3
C1.3	Number of years back from base (to match historical data)	0	1	2	3	4
C1.4	Found rate (cumulative entrants found in range of years as a proportion of total found in 2017)	0 (no matching with historical records)	20/70 = 0.29	13/70 = 0.19	8/70 = 0.11	3/70 = 0.04

Once found rates are obtained, the number of students found for previous years can be adjusted accordingly in order to account for the fewer years of historical data available. For example, in example C2, matching student records for 2016 to previous years' records only allows for 3 years of historical records instead of 4 years for when records for 2017 were matched. The found rate for searching 4 years of historical records are multiplied by the actual number of total entrants in 2016 to obtain the proportion of students who should have been found. In example C2, this number is 2, suggesting that an additional 2 people should be categorized as a previous entrant instead of as a new entrant. This number is then subtracted from the number of new entrants for the year and added to the number of found entrants, resulting in 28 new entrants instead of 30, and 22 previous entrants instead of 20. This adjustment occurs for all years prior to the current year in order to adjust the number of new and previous higher education entrants accordingly.

Example C2: Adjusting found rates for 2016

Year (current = 2016)	2015	2014	2013	2012
Number of years back	1	2	3	4
Actual number of students found	8	7	5	N/A
Actual found rate for 2016	0.4	0.35	0.25	N/A
Applied found rate (from example C1)	-	-	-	0.04
Extra found (applied found * total entrants)	-	-	-	$0.04 * 50 = 2$
Adjusted number of students found	8	7	5	2

Accounting for historical data 2016		
	Number	Proportion
Total entrants	50	-
Unadjusted new entrants	30	$30/50$ = 0.6
Unadjusted found entrants	$8+7+5$ = 20	$20/50$ = 0.4
Adjusted new entrants	$30 - 2$ = 28	$28/50$ = 0.56
Adjusted found entrants	$8+7+5+2$ = 22	$22/50$ = 0.44

After adjusting populations to reflect the expected proportion of the population who have previously participated in higher education, the number of new entrants in the particular year is obtained by counting all students whose records did not match with any historical records. This number is used to calculate the rate of new entrants in higher education in each year, categorised by age and quintile. The rate of new entrants is calculated as the number of new entrants as a proportion of the population (adjusted) for each age group and quintile.

Annex D: Participation by age and POLAR4 quintile

Obtaining the rate of new entrants for each previous year, grouped by age and quintiles, allows for the identification of the proportion of new students in each year. The sum of all these rates would then represent the proportion of people who have previously attended higher education for each particular age group and quintile.

The year which the calculation of rates ends will depend on the age group because the calculation has to go back to the year in which they were 18. For example, if participation rates for 2016 for the 30 age group were to be calculated, the new entrants for previous years would date back 12 years to 2002. Conversely, for participation rates in 2016 for the 18 age group, no previous years of new entrants are calculated, since it is the first year possible for the age group to have entered higher education.

One key point to note is that while the overall participation rate for a particular group is the sum of new entrant rates for each year, the group's age changes from year to year. For example, participation for the aged 30/quintile 1 group for 2016 (current year) would be the new entrant rates for the group in 2016 combined with the new entrant rates for previous years. However, the age of the aged 30/quintile 1 group for previous years will not be 30. Instead, in 2015 the cohort would be aged 29/quintile 1, for 2014 aged 28/quintile 1, for 2013

aged 27/quintile 1, etc. With this in mind, the participation rate for each group can be calculated.

$$P_{(y,a,p)} = R_{(y,a,p)} + R_{(y-1,a-1,p)} + R_{(y-2,a-2,p)} + \cdots + R_{(y-n,a-n,p)}$$

Where: P = Participation Rate

R = Rate of new entrants for each group

y = current year

a = age group

p = POLAR4 quintile

n = number of years of previous new entrants necessary (a-18)

For example, the participation rate for the group aged 30 from quintile 1 in 2016:

$$P_{(2016,30,1)} = R_{(2016,30,1)} + R_{(2015,29,1)} + R_{(2014,28,1)} + \cdots + R_{(2004,18,1)}$$

The rates calculated are the proportions for each specific age group and quintile. This allows us to have the participation rate for each combination of ages (18-30) and quintiles (1-5). Example D1 demonstrates an example of these calculations for just quintile 1, assuming that we are calculating rates only for those aged 18-20 instead of the full 18-30 in 2016. In this example, the rate for new entrants in higher education in 2016 for 18 year olds is 0.2, and because 18 year olds are not counted in higher education before they turned 18, the participation rate for the Q1/18 year old group is 0.2. For the 19 year old age group, 0.1 of the 19 year old population entered higher education for the first time in 2016, while 0.18 entered the year prior when they were 18. This means the participation rate for the Q1/19 year old group is 0.1 + 0.18, which amounts to 0.28. Similarly, the participation rate for the Q1/20 year old group consists of the rates of their entrant to higher education in past years, which is 0.34. While this is only an example, in reality the participation rates are calculated in the same manner, but the groups range from ages 18-30 for each POLAR4 quintiles – 60 groups in total with its own participation rate.

Example D1:

POLAR4	Age	Participation Rate	Rate 2016	Rate 2015	Rate 2014
Q1	18	0.2	0.20	0	0
	19	0.28	0.10	0.18	0
	20	0.34	0.06	0.12	0.16

However, because this rate is still grouped by age and quintile it does not represent the overall participation rate for each POLAR4 quintile. In order to obtain rates which measure participation in higher education for each quintile, the participation rates for each age (proportion of a particular age group who have been to or are currently in higher education) are multiplied by their respective populations for each age in order to obtain the total number of people in that age group who have been to or are currently in higher education. Participation rates for each age group have been calculated separately in order to weigh the final participation rate for each of the quintiles by the population of each age group and not assume that age is evenly distributed along the whole population. After multiplying the population of each age group by its participation rate, the sum of these for each age group would then result in the total number of 18-30 year olds who have been to or are currently in higher education for each POLAR4 quintile.

Example D2:

POLAR4	Age	Participation Rate	Population	Total number of participants
Q1	18	0.2	100	$0.2 \times 100 = 20$
	19	0.28	120	$0.28 \times 120 = 33.6$
	20	0.34	90	$0.34 \times 90 = 30.6$
Total				$20 + 33.6 + 30.6 = 84.2$

Example D2 displays the process of how the total number of people who participate in higher education for the current year in each POLAR4 quintile is calculated. In the example, it is assumed that we are only counting students aged 18-20. As can be seen, the participation rates derived in example D1 are multiplied by their respective populations, which is equal to

the total number of higher education participants of that particular age in quintile 1. The sum of this is the total number of participants in quintile 1 in 2016 (assuming it is the current year).

The number of total participants, grouped by POLAR4 quintiles, is divided by the adjusted population count for each quintile. The same calculations are repeated for each quintile, resulting in the overall higher education participation rate grouped by POLAR4 quintiles (see example D3).

Example D3:

POLAR4	Number of higher education participants	Adjusted population	Final participation rate
Q1	84.2	500	16.84%
Q2	90.6	480	18.875%
Q3	110.2	550	20.04%
Q4	116.1	450	25.80%
Q5	134.3	500	26.86%

Annex E: List of high tariff institutions

The classification of institutions by entry tariff levels includes higher education institutions and excludes specialist higher education institutions, further education colleges, and alternative higher education providers. It is based on HESA student data from the academic years 2012-13 to 2014-15.

English higher education institutions (HEIs) have been ordered by the average tariff score of their total young (aged under 21) UK-domiciled undergraduate entrants in the 2012-13 to 2014-15 academic years. The average tariff score calculation considers all such entrants holding level 3 qualifications which are subject to the UCAS Tariff. For each higher education institution, the 5 per cent tails of the tariff score distribution have been removed. Higher education institutions in the top third of the ranking by average tariff score form the "HEIs with high average tariff scores" group, and those in the bottom third comprise the "HEIs with low average tariff scores" group.

Higher education institution (HEI) with high tariff scores

Aston University

Imperial College London

King's College London

Loughborough University

Queen Mary University of London

Royal Holloway, University of London

The London School of Economics and Political Science

The School of Oriental and African Studies

The University of Bath

The University of Birmingham

The University of East Anglia

The University of Lancaster

The University of Leeds

The University of Leicester

The University of Liverpool

The University of Manchester

The University of Nottingham

The University of Reading

The University of Sheffield

The University of Surrey

The University of Warwick

University College London

University of Bristol

University of Cambridge

University of Durham
University of Exeter
University of Newcastle upon Tyne
University of Oxford
University of Southampton
University of Sussex
University of York